

A joint model of time to breast cancer recurrence and CA15-3

Cite as: AIP Conference Proceedings **2040**, 110002 (2018); <https://doi.org/10.1063/1.5079166>
Published Online: 30 November 2018

Ana Borges, Inês Sousa, and Luis Castro



View Online



Export Citation

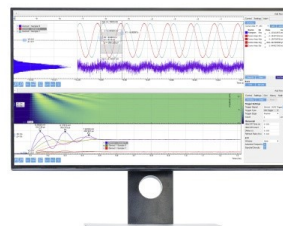
ARTICLES YOU MAY BE INTERESTED IN

[Joint modelling of longitudinal CEA tumour marker progression and survival data on breast cancer](#)

AIP Conference Proceedings **1836**, 020043 (2017); <https://doi.org/10.1063/1.4981983>

Challenge us.

What are your needs for
periodic signal detection?



Zurich
Instruments

A Joint Model of Time to Breast Cancer Recurrence and CA15-3

Ana Borges^{1,a)}, Inês Sousa^{2,b)} and Luis Castro^{3,c)}

¹*CIICESI/ESTG - P.Porto, Margaride 4610-156 Felgueiras, Portugal.*

²*CMAT, DMA-ECUM, Universidade do Minho, Campus de Gualtar 4710 - 057 Braga, Portugal.*

³*Hospital de Braga, 4710-243 Braga, Portugal.*

^{a)}Corresponding author: aib@estg.ipp.pt

^{b)}isousa@math.uminho.pt

^{c)}luismedicina@yahoo.com

Abstract. The main purpose of the present analysis is to jointly analyse longitudinal data of the tumour marker Carcinoma Antigen 15-3 (CA15-3), and the probability of recurrence of patients with breast cancer, by incorporating a spatial effect in order to infer how patients' residence location affects the progression of the disease. A data set of 540 patients was collected from medical records of Braga's Hospital, in Portugal. A joint model of the two stochastic processes was used, as it is expected an association between longitudinal progression of the tumour marker and time-to-recurrence. Results suggest that the longitudinal progression of CA15-3 is significantly associated with the probability of breast cancer recurrence for these patients. Factors such as images of vascular invasion affect the longitudinal process. While factors such as triple negative tumour, primary tumour size and Regional Lymph nodes involvement degree affect the probability of recurrence. The spatial effect is significant on both longitudinal and time-to-recurrence process.

INTRODUCTION

Breast cancer is an extremely complex disease characterized by numerous factors that can affect its progression in an unexpected way. Currently, only a few studies characterizing the complexity of this cancer on Portuguese population exist, and are mainly conducted by population-based cancer registries, as for example RORENO (North Regional Cancer Registry). Also, the study from the ROR-SUL Working Group [1] is worthy of pointing out as it contributes to detect survival differences in female breast cancer and identify the main associated factors, such as geographical variation, on the south region of Portugal. Knowing which risk factors affect a certain population diagnosed with breast cancer, based on both cause-specific death information and recurrence information, can improve the knowledge of its progression, probability of recurrence or even survival pattern and, undeniably, lead to better treatment and prognostic. In this sense, the likelihood that the patient will suffer a recurrence of her disease is very important, so that the risks and expected benefits of specific therapies can be compared [7].

In view of the above, along with the scarcity of studies concerning the progression of the disease specifically for the Portuguese population, particularly with regard to relapse, justifies the motivation for the present work. Our main objective is, therefore, to identify risk factors that affect the probability of breast cancer relapse for patients diagnosed with breast cancer and treated in Braga's Hospital, located in north of Portugal. Aiming to contribute to the understanding of this complex disease, in the Portuguese context. We study two response variables commonly measured in health sciences: a longitudinal variable, which is a variable repeatedly measured in time for different subjects; and a time from an origin up to an event of interest. Particularly, we analyse a longitudinal variable: the Carcinoma Antigen 15-3 (CA15-3), which consists in a large roll of measures obtained in all blood test of each patient. Meaning that we have several CA15-3 values and the respective moments of their measurements. Also, we are interested in studying the time from breast cancer diagnose until breast cancer recurrence.

Biomarkers of cancer are blood-based molecules that indicate malignant properties, allowing for minimally invasive testing, which may help in diagnosing the disease at an early stage, to categorize tumours to select an appropriate

treatment, to monitor the response to therapy, or to assess disease progression and recurrence, as in our particular case [9]. CA15-3 is the most widely used serum marker in breast cancer. The main uses of CA 15-3 are in preclinically detecting recurrent breast cancer and monitoring the treatment of patients with advanced breast cancer [5].

From previous analysis [4] we found a significant association between the progression in time of CA15-3 and the probability of dying, from breast cancer, as so, we adopt the methodology of random effects [6, 10] to adjust a joint model of the longitudinal variable and the time-to-relapse, in order to infer on the association between them, and also to understand which factors affect each. Also, to understand if geographical variation is related to the probability of breast cancer recurrence and the progression of the CA15-3 tumour marker it was tested a spatial covariate.

The analyzed data gathers information on patients of the Senology Unit of Braga's Hospital, located in the north of Portugal, diagnosed with breast cancer. In this work, we begin by briefly describing the data set collected from Braga Hospital records. Later we clarify the statistical methodology implemented on the joint analysis of the longitudinal and time-to-event processes. Then we expose the main results, ending with a discussion section.

Methodology

For this study we work with data collected directly from the hospital medical records of each patient diagnosed with breast cancer, listed in the computer system of Braga's Hospital. The authorization for the collection and use of data Senology was approved by Braga's Hospital Ethics Committee. From the information gathered in all the medical reports we were able to collect more than 50 variables. They can be grouped into two categories: (i) explanatory variables at individual level, which are a group of demographic characteristics, prognostic and etiologic factors (i.e. a cause or origin of a disease) as for example: age, menopause, family history of breast cancer, etc.; and (ii) explanatory variables at tumour level, that include characteristics of the tumour, such as TNM stage, histological type of tumour, hormonal receptors or vascular or lymphatic invasion, among others [8].

All variables that were significant in previous analysis [2, 3, 4], were tested both in the longitudinal process and the time-to-event process. To account for geographical variation, a covariate was created based on the Classification of urban areas (TIPAU 2014). This classification was created for statistical purposes, as explained by Portuguese Statistical Institute (INE), and consists in a triple classification of the national parishes in: (i) predominantly urban areas (APU); (ii) medium urban areas (AMU), and (iii) predominantly rural areas (APR). Hence, the spatial covariate named TIPAU consists in those three categories.

The present joint analysis adopts of random effect methodology developed by Wulfsohn and Tsiatis [10] with the extension of Henderson [6], making use of a packages of R software: *JM* [11].

The model implemented uses a linear mixed effects model, to represent the longitudinal process, given by:

$$\text{Longitudinal Process: } Y_{ij} = \mu_{ij} + U_i + Z_{ij} = X_1\beta + U_{0i} + U_{1i}t_{ij} + Z_{ij}, \quad (1)$$

where each patient is denoted by the subscript $i = 1, \dots, n$, and repeated measurements for each patient i , at corresponding time t_{ij} , are denoted by Y_{ij} . X_1 is the design matrix for the fixed covariates, with corresponding regression parameters β . The latent variable given by (U_{0i}, U_{1i}) is a realization of a $MVN(0, \Sigma)$ where $\Sigma = \begin{pmatrix} v_1^2 & v_{12} \\ v_{12} & v_2^2 \end{pmatrix}$ is the variance/covariance matrix of the random effects. Z_{ij} are Normal i.i.d. realizations of $N(0, \tau^2)$, representing the measurement error (variability non specified).

JM incorporates the precise estimate of the longitudinal unobserved process, $m_i(t) = \hat{\mu}_{ij} + \hat{U}_i$, into the survival process as it follows:

$$\text{JM Survival Process: } h_i(t) = h_0(t) \exp\{X_{2i}\beta_2^* + \alpha m_i(t)\}, \quad (2)$$

where X_{2i} represents the baseline covariates, and β_2^* the vector of corresponding regression parameters. The parameter α represents the effect of the true longitudinal response on the survival process. The correspondent parameters in β_2^* represent the extra effect in the survival sub-model of the baseline covariate. To synthesize, we will be able to determinate the factors that influence the change in CA15-3 values, on the patient's and what effect that change has on their survival.

We consider the beginning of the study as the date of the malignant tumour diagnosis and the end of the study the date 30/11/2014. In this particular analysis we know the moment that the individuals enter the study, as well the date of the recurrence if it occurs until the end of the study. However, as we do not know the precise moment of the recurrence if it occurs after the end of the study, we have to take into account a right censoring of the data. For model

diagnose we use graphical representation of the subject specific residuals versus the fitted responses for the validity of the mixed effects model. This plot is used to assess the assumption of constant variance of Z_{ij} . Also, a Q-Q plot of the subject-specific residuals is represented to verify the assumption of normality of Z_{ij} . To validate the fitted survival process we made use of the graphical representation of Cox-Snell residuals.

Main Results

Information was gathered on 540 patients for the present analysis. As 19 patients had cancer of bilateral breast, and by suggestion of the consultant doctor of this study, they were treated as separate cases, which resulted in a total number of 559 cases analyzed in female patients diagnosed with malignant tumor from 2008 to 2012. The total number of deaths was 74, however, the total number of deaths from breast cancer is only 55. Throughout the followup time it was detected a recurrence of breast cancer (local and/or distant) on 81 patients. Of these, 52 died from breast cancer. We accounted for a majority of patients (55.35%) residing in predominantly urban areas which, also, represents more than half of the deaths from breast cancer on this population (53.33%). Only patients with information of CA15-3 values available were included. Which means that only 534 patients were eligible for the present analysis. With a total number of tumour marker measurements of 5166. The normality assumption of the response variable CA15-3 was rejected, by performing a Shapiro-Wilk normality test, with returned p-value < 0.0001 . Hence, it was used a log-transformation of tumor markers values.

Table 1 presents the results from joint modelling the longitudinal process of CA15-3 tumour marker values and the time-to-recurrence process, where the reference time is the time (in months) since tumour diagnose until blood test data. It is shown the estimates obtained for the covariates considered in the longitudinal and survival process, as well its standard errors (SE) and respective p-values.

TABLE 1. Estimated parameters values for Joint Model - CA15-3 tumour marker

Covariates	Estimates	SE	p-value		Estimates	SE	p-value
LONGITUDINAL PROCESS				LATENT ASSOCIATION			
Intercept	2.504	0.145	< 0.0001	Association parameter α	1.272	0.124	< 0.0001
Time (month)	0.0002	0.001	0.681	ν_1^2		0.4102	
Age at diagnosis	0.003	0.003	0.239	ν_2^2		0.00001	
Images of vascular invasion(Yes)	1.399	0.119	< 0.0001	τ^2		0.2121	
TIPAU(APR)	-0.008	0.104	0.941				
TIPAU(APU)	-0.748	0.102	< 0.0001				
SURVIVAL PROCESS							
Triple Negative (with)	3.238	0.502	< 0.0001				
Tumour size (T2)	-0.286	0.486	0.557				
Tumour size (T3)	1.210	0.505	0.017				
Tumour size (T4)	6.460	1.491	0.998				
Tumour size (Tis)	2.056	1.06	0.053				
Tumour size (Tx)	2.945	0.736	< 0.0001				
Regional Lymph nodes (NX or N0 or N1)	-1.135	0.378	0.002				
TIPAU(APR)	-1.535	0.561	0.006				
TIPAU(APU)	-0.9023	1.069	0.398				

It is shown that covariates such as time and age at diagnosis do not have a significant effect, despite the results obtained on previous individual longitudinal studies [4]. The presence of images of vascular invasion and TIPAU have significant effect on initial values of the linear mean progression of CA15-3 log value. For instance, a tumour with images of vascular invasion presents an increment of 1.399 on the starting value of the tumour marker linear progression. An important result is that for a patient residing in predominantly urban areas (APU) causes a decrease of 0.748 on initial values of the linear mean progression of CA15-3 log values, compared to patients residing in medium urban areas (AMU). Also, there is no significant difference on the starting values of the linear progression of the tumour marker between patients from medium urban areas (AMU) and from predominantly rural areas (APR).

Concerning the survival process, covariates Triple Negative, primary tumour size, regional lymph nodes involvement degree and TIPAU have a significant effect on the risk of breast cancer recurrence. As expected, from results already reported in earlier studies, a triple negative type of tumour has a higher risk of dying from breast cancer.

In terms of the spatial covariate, results indicate that patients from predominantly rural areas (APR) areas have a lower probability of breast cancer recurrence, with an hazard ratio of 0.215 ($\exp(-1.535)$), compared to patients residing in medium urban areas (AMU). Also, there is no significant difference on survival between patients from medium urban areas (AMU) and patients residing in predominantly urban areas (APU). The significance of the association parameter $\hat{\alpha}$ ($p\text{-value} < 0.0001$) in the survival process highlights the importance for a joint model for the present data. This significant association verifies the relationship between the longitudinal and survival processes. Through determination of hazard ratio (HR), it is apparent that individuals with higher CA15-3 log value tend to have a worst survival ($HR = \exp(1.272) \approx 3.57$). This was expected as high values of CA15-3 corresponds to a deterioration of disease.

Discussion

The work here presented aims to identify risk factors that affect the probability of breast cancer relapse for patients diagnosed with breast cancer from Braga's Hospital, located in north of Portugal. Results show that longitudinal CA15-3 values are significantly associated with the probability of breast cancer recurrence. Where higher values of the tumour marker are related to higher risk of breast cancer recurrence. Modelling together the survival and longitudinal processes shows that the association between both processes is significant and it is essential its recognition. Concerning the probability of breast cancer recurrence, it was found that covariates Triple Negative, primary tumour size, regional lymph nodes involvement degree and TIPAU have a significant effect. The spatial effect is significant on both longitudinal CA15-3 and survival processes. Also, no significant differences were register between the probability of breast cancer recurrence of patients from medium urban areas (AMU) and predominantly urban areas (APU). It was detected that, the risk of breast cancer recurrence is significantly lower for a patient from a predominantly rural area (APR) compared to a patient who resides in an AMU area. The results point to a possible geographical variation in breast cancer epidemiology, contributing to the understanding of this disease in the Portuguese population.

ACKNOWLEDGMENTS

Ana Borges thanks the Fundação para a Ciência e Tecnologia (FCT), Portugal, for the Ph.D. Grant SFRH/BD/74166/2010, and the support by the Center for Research and Innovation in Business Sciences and Information Systems(CIICESI), ESTG - P.Porto.

REFERENCES

- [1] do Rosario Andr, M., Amaral, S., Mayer, A., Miranda, A. and Working Group ROR SUL. (2014). *Acta medica portuguesa*, 27(3), 325-330.
- [2] Borges, A., Sousa, I. and Castro, L. (2013). *Book of Abstracts-ASMDA2013*, 34-35.
- [3] Borges, A., Sousa, I. and Castro, L. (2015). *REVSTAT Statistical Journal*, 13, 63-78.
- [4] Borges, A. I. Sousa, I. and Castro, L. (2016). *Atas do XXII Congresso da Sociedade Portuguesa de Estatística.*, 27 71. ISBN: 978-972-8890-39-1.
- [5] Duffy, M. J., Shering, S., Sherry, F., McDermott, E. and O'higgins, N. (2000). *The International journal of biological markers*, 15(4), 330-333.
- [6] Henderson, R., Diggle, P. and Dobson, A. (2000). *Biostatistics*, 1(4), 465-480.
- [7] Jerez-Aragonés, J. M., Gómez-Ruiz, J. A., Ramos-Jiménez, G., Munoz-Pérez, J. and Alba-Conejo, E. (2003). *Artificial intelligence in medicine*, 27(1), 45-63.
- [8] Rodrigues, V. (2011), Chapter 34. In *Manual de Ginecologia*, Permanyer, Portugal, 175-191.
- [9] Shankar, P. S. (2017). *RGUHS Journal of Medical Sciences*, 7(2), 49-50.
- [10] Wulfsohn, M. S. and Tsiatis, A. A. (1997). *Biometrics*, 330-339.
- [11] Rizopoulos, D. (2010). *Journal of Statistical Software (Online)*, 35(9), 1-33.